



## Unifying Markov properties for graphical models

Lauritzen, Steffen L.; Sadeghi, Kayvan Sadeghi

*Published in:*  
Annals of Statistics

*DOI:*  
[10.1214/17-AOS1618](https://doi.org/10.1214/17-AOS1618)

*Publication date:*  
2018

*Document version*  
Publisher's PDF, also known as Version of record

*Citation for published version (APA):*  
Lauritzen, S. L., & Sadeghi, K. S. (2018). Unifying Markov properties for graphical models. *Annals of Statistics*, 46(5), 2251-2278. <https://doi.org/10.1214/17-AOS1618>

# UNIFYING MARKOV PROPERTIES FOR GRAPHICAL MODELS

BY STEFFEN LAURITZEN AND KAYVAN SADEGHI<sup>1</sup>

*University of Copenhagen and University of Cambridge*

Several types of graphs with different conditional independence interpretations—also known as Markov properties—have been proposed and used in graphical models. In this paper, we unify these Markov properties by introducing a class of graphs with four types of edges—lines, arrows, arcs and dotted lines—and a single separation criterion. We show that independence structures defined by this class specialize to each of the previously defined cases, when suitable subclasses of graphs are considered. In addition, we define a pairwise Markov property for the subclass of chain mixed graphs, which includes chain graphs with the LWF interpretation, as well as summary graphs (and consequently ancestral graphs). We prove the equivalence of this pairwise Markov property to the global Markov property for compositional graphoid independence models.

**1. Introduction.** Graphical models provide a strong and clear formalism for studying conditional independence relations that arise in different statistical contexts. Originally, graphs with a single type of edge were used; see, for example, [3] for undirected graphs (originating from statistical physics [11]), and [13, 40] for directed acyclic graphs (originating from genetics [43]).

With the introduction of chain graphs [18], and other types of graphs with edges of several types [2, 23, 26, 39] as well as different interpretations of chain graphs [1, 6], a plethora of Markov properties have emerged. These have been introduced with different motivations: chain graphs as a unification of directed and undirected graphs, the so-called AMP Markov property to describe dependence structures among regression residuals, bidirected graphs to represent structures of marginal independence and other mixed graphs to represent selection effects and incomplete observations in causal models. Despite the similarities among these, the lack of a general theory as well as the use of different definitions and notation has undermined the original conceptual simplicity of graphical models. This motivates a unification of the corresponding Markov properties. In [29], we attempted this for different types of mixed graphs, but failed to include chain graph Markov properties. Here, we follow an analogous approach using a single separation criterion,

---

Received August 2016; revised July 2017.

<sup>1</sup>Supported in part by Grant #FA9550-14-1-0141 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

*MSC2010 subject classifications.* Primary 62H99; secondary 62A99.

*Key words and phrases.* AMP Markov property, *c*-separation, chain graph, compositional graphoid, *d*-separation, independence model, LWF Markov property, *m*-separation, mixed graph, pairwise Markov property, regression chain Markov property.

but using four types of edges: line, arrow, arc and dotted line. To the best of our knowledge, this unifies most graphical independence models previously discussed in the literature. One exception is Drton's [6] type III chain graph Markov property, which has several unfortunate properties and so far has not played any specific role; we have chosen to avoid introducing a fifth type of edge to accommodate this property; another exception is the *reciprocal* graphs of Koster [14], which allow feedback cycles; other exceptions use graphs to describe conditional independence in dynamical systems [5, 8], which we do not discuss here. Our unification includes summary graphs—which include ancestral graphs as well as chain graphs with the multivariate regression Markov property [2]—chain graphs with the LWF Markov property [9, 18], and chain graphs with the AMP Markov property [1].

In addition to the unification of the (global) Markov property, we provide a unified pairwise Markov property. However, it seems technically complex to include the pairwise Markov property for chain graphs with the AMP interpretation, and hence we only discuss this for the subclass of graphs with three types of edges where cycles of specific types are absent. Such graphs were called chain mixed graphs (CMGs) in [28] and its corresponding independence model unifies those of summary graphs (and ancestral graphs) as well as chain graphs with the LWF Markov property. For CMGs, we first discuss the notion of maximality and show that every missing edge in a maximal CMG corresponds to an independence statement, thus forming a potential base for specifying pairwise Markov properties. For CMGs, we prove the equivalence of pairwise and global Markov properties for abstract independence models which are compositional graphoids.

The structure of the paper is as follows: In the next section, we define graphs with four types of edges and provide basic graph theoretical definitions. In Section 3, we discuss general independence models and compositional graphoids, provide a single separation criterion for such graphs, and show that the induced independence models are compositional graphoids. Further we demonstrate how the various independence models discussed in the literature are represented within this unification. In Section 4, we define the notion of maximal graphs, provide conditions under which a CMG is maximal and show that any CMG can be modified to become maximal without changing its independence model. In Section 5, we provide a pairwise Markov property for CMGs, and prove that for compositional graphoids, the pairwise Markov property is equivalent to the global Markov property. Finally, we conclude the paper with a discussion in Section 6.

## 2. Graph terminology.

**2.1. Graphs.** A graph  $G$  is a triple consisting of a *node* set or *vertex* set  $V$ , an *edge* set  $E$ , and a relation that with each edge associates two nodes (not necessarily distinct), called its *endpoints*. When nodes  $i$  and  $j$  are the endpoints of an edge, these are *adjacent* and we write  $i \sim j$ . We say the edge is *between* its two endpoints. We usually refer to a graph as an ordered pair  $G = (V, E)$ . Graphs

$G_1 = (V_1, E_1)$  and  $G_2 = (V_2, E_2)$  are called *equal* if  $(V_1, E_1) = (V_2, E_2)$ . In this case, we write  $G_1 = G_2$ .

The graphs that we use are *labeled graphs*, that is, every node is considered a different object. Hence, for example, the graph  $i \text{ --- } j \text{ --- } k$  is not equal to the graph  $j \text{ --- } i \text{ --- } k$ .

In addition, in this paper, we use graphs with four types of edges denoted by *arrows*, *arcs* (solid lines with two-headed arrows), *lines* (solid lines) and *dotted lines*; as will be seen in Section 3, we shall use dotted lines to represent chain graphs with the AMP Markov property. Henceforth, by ‘graph’, we mean a graph with these four possible types of edges. We do not distinguish between  $i \text{ --- } j$  and  $j \text{ --- } i$ , between  $i \longleftrightarrow j$  and  $j \longleftrightarrow i$ , or between  $i \cdots \cdots j$  and  $j \cdots \cdots i$ , but we do distinguish between  $j \rightarrow i$  and  $i \rightarrow j$ .

A *loop* is an edge with endpoints being identical. In this paper, we are only considering graphs that do not contain loops. *Multiple edges* are edges sharing the same pair of endpoints. A *simple graph* has neither loops nor multiple edges. Graphs we are considering in this paper may generally contain multiple edges, even of the same type. However, we shall emphasize that for all purposes in the present paper, multiple edges of the same type are redundant, and hence at most one edge of every type is necessary to represent the objects we discuss.

We say that  $i$  is a *neighbor* of  $j$  if they are endpoints of a line; if there is an arrow from  $i$  to  $j$ ,  $i$  is a *parent* of  $j$  and  $j$  is a *child* of  $i$ . We also say that  $i$  is a *spouse* of  $j$  if they are endpoints of an arc, and  $i$  is a *partner* of  $j$  if they are endpoints of a dotted line. We use the notation  $\text{ne}(j)$ ,  $\text{pa}(j)$ ,  $\text{sp}(j)$  and  $\text{pt}(j)$  for the set of all neighbours, parents, spouses and partners of  $j$ , respectively. More generally, for a set of nodes  $A$  we let  $\text{ne}(A) = \bigcup_{j \in A} \text{ne}(j) \setminus A$  and similarly for  $\text{pa}(A)$ ,  $\text{sp}(A)$  and  $\text{pt}(A)$ .

A *subgraph* of a graph  $G_1$  is a graph  $G_2$  such that  $V(G_2) \subseteq V(G_1)$  and each edge present in  $G_2$  also occurs in  $G_1$  and has the same type there. An *induced subgraph* by a subset  $A$  of the node set is a subgraph that contains all and only nodes in  $A$  and all edges between two nodes in  $A$ .

A *walk*  $\omega$  is a list  $\omega = \langle i_0, e_1, i_1, \dots, e_n, i_n \rangle$  of nodes and edges such that for  $1 \leq m \leq n$ , the edge  $e_m$  has endpoints  $i_{m-1}$  and  $i_m$ . We allow a walk to consist of a single node  $i_0 = i_n$ . If the graph is simple, then a walk can be determined uniquely by a sequence of nodes. Also, a nontrivial walk is always determined by its edges, so we may write  $\omega = \langle e_1, \dots, e_n \rangle$  without ambiguity. Throughout this paper, however, we often use only node sequences to describe walks even in graphs with multiple edges, when it is apparent from the context or the type of the walk which edges are involved. The first and the last nodes of a walk are its *endpoints*. All other nodes are *inner nodes* of the walk. We say a walk is *between* its endpoints. A *cycle* is a walk with at least two edges and no repeated node except  $i_0 = i_n$ . A *path* is a walk with no repeated node.

A *subwalk* of a walk  $\omega = \langle i_0, e_1, i_1, \dots, e_n, i_n \rangle$  is a walk that is a subsequence  $\langle i_r, e_{r+1}, i_{r+1}, \dots, e_p, i_p \rangle$  of  $\omega$  between two occurrences of nodes  $(i_r, i_p, 0 \leq r \leq p \leq n)$ . If a subwalk forms a path, then it is also a *subpath* of  $\omega$ .

In this paper, we need different types of walks as defined below. Consider a walk  $\omega = \langle i = i_0, i_1, \dots, i_n = j \rangle$ . We say that:

- $\omega$  is *undirected* if it only consists of solid lines;
- $\omega$  is *directed* from  $i$  to  $j$  if all edges  $i_q i_{q+1}$ ,  $0 \leq q \leq n-1$ , are arrows pointing from  $i_q$  to  $i_{q+1}$ ;
- $\omega$  is *semi-directed* from  $i$  to  $j$  if it has at least one arrow, no arcs and every arrow  $i_q i_{q+1}$  is pointing from  $i_q$  to  $i_{q+1}$ ;
- $\omega$  is *anterior* from  $i$  to  $j$  if it is semi-directed from  $i$  to  $j$  or if it is composed of lines and dotted lines.

Thus a directed walk is also semi-directed and a semi-directed walk is also an anterior walk. If there is a directed walk from  $i$  to  $j$  ( $j \neq i$ ), then  $i$  is an *ancestor* of  $j$ . We denote the set of ancestors of  $j$  by  $\text{an}(j)$ . If there is an anterior walk from  $i$  to  $j$  ( $j \neq i$ ), then we also say that  $i$  is *anterior* of  $j$ . We use the notation  $\text{ant}(j)$  for the set of all antérieurs of  $j$ . For a set  $A$ , we define  $\text{ant}(A) = \bigcup_{i \in A} \text{ant}(i) \setminus A$ . We also use the notation  $\text{An}(A)$  and  $\text{Ant}(A)$  for the set of *reflexive* ancestors and antérieurs of  $A$  so that  $\text{An}(A) = A \cup \text{an}(A)$  and  $\text{Ant}(A) = A \cup \text{ant}(A)$ . In addition, we define a set  $A$  to be *anterior* if  $\text{ant}(i) \subseteq A$  for all  $i \in A$ ; in other words,  $A$  is anterior if  $\text{ant}(A) = \emptyset$ .

In fact, we are only interested in these walks when we discuss graphs without dotted lines. For example, consider the following walk (path) in such a graph:

$$i \text{ --- } j \text{ --- } k \longrightarrow l \longrightarrow m \text{ --- } n \longrightarrow o \longleftrightarrow p.$$

Here, it holds that there is an undirected walk between  $i$  and  $k$ , and hence  $i \in \text{ant}(k)$ , but there is no semi-directed walk from  $i$  to  $k$ . In addition, we have that  $k \in \text{an}(m)$  and  $i \in \text{ant}(o)$ , while there is a semi-directed walk from  $i$  to  $o$ . There is also no anterior walk from  $i$  to  $p$ .

Notice that, unlike most places in the literature (e.g., [26]), we use walks instead of paths to define ancestors and antérieurs. Using walks instead of paths is immaterial for this purpose as the following lemma shows.

**LEMMA 1.** *There is a directed or anterior walk from  $i$  to  $j$  if and only if there is a directed or anterior path from  $i$  to  $j$ , respectively.*

**PROOF.** If there is a path, there is a walk as a path is also a walk. Conversely, assume there is a directed or anterior walk from  $i$  to  $j$ . If  $i = j$ , then we are done by definition. Otherwise, start from  $i$  and move on the walk towards  $j$ . Consider the first place where a node  $k$  is repeated on the walk. The walk from  $k$  to  $k$  forms a cycle. If we remove this cycle from the walk, the resulting walk remains directed; similarly, the walk resulting from an anterior walk remains anterior. Successively removing all cycles along the walk in this way implies the result.  $\square$

A *section*  $\rho$  of a walk is a maximal subwalk consisting only of solid lines, meaning that there is no other subwalk that only consists of solid lines and includes  $\rho$ .

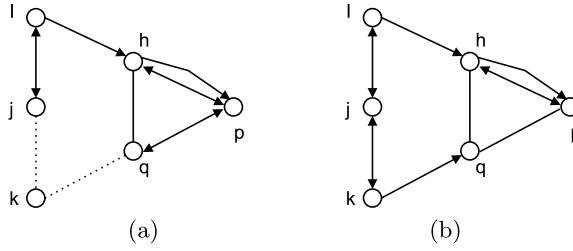


FIG. 1. (a) A graph with four types of edges. (b) A graph that is not a CMG as  $\langle h \rightarrow p \rightarrow q \rightarrow h \rangle$  is a semi-directed cycle.

A walk decomposes uniquely into sections; sections may also be single nodes. The section is an *inner section* on the walk if all nodes on the section are inner nodes on the walk and an *endpoint section* if it contains an endpoint of the walk. A section  $\rho$  on a walk  $\omega$  is called a *collider section* if one of the four following walks is a subwalk of  $\omega$ :  $u \rightarrow \rho \leftarrow v$ ,  $u \leftrightarrow \rho \leftarrow v$ ,  $u \leftrightarrow \rho \leftrightarrow v$ ,  $u \rightarrow \rho \cdots v$  and  $u \leftrightarrow \rho \cdots v$ , that is, a section  $\rho$  is a collider if two arrowheads meet at  $\rho$  or an arrowhead meets a dotted line. All other sections on  $\omega$  are called *noncollider* sections; these are sections that are an endpoint of  $\omega$  or the following sections:  $u \leftarrow \rho \rightarrow v$ ,  $u \leftarrow \rho \leftrightarrow v$ ,  $u \leftarrow \rho \cdots v$ ,  $u \rightarrow \rho \rightarrow v$  and  $u \cdots \rho \cdots v$ . We may speak of collider or noncollider sections (or nodes) without mentioning the relevant walk when this is apparent from the context. Notice that a section may be a collider on one part of the walk and a noncollider on another. For example, in Figure 1(a), the section  $\langle h, q \rangle$  is a collider on the walk  $\langle l, h, q, p \rangle$ . It is also a collider on  $\langle k, q, h, p \rangle$  via the edge  $h \leftrightarrow p$ , but a noncollider on  $\langle k, q, h, p \rangle$  via the edge  $h \rightarrow p$ . Notice also that  $\langle k \rangle$  is a noncollider on  $\langle j, k, q \rangle$ .

A *tripath* is a path with three distinct nodes. Note that [27] used the term V-configuration for such a path. If the inner node on a tripath is a collider, we shall also say that the tripath itself is a collider or noncollider.

**2.2. Subclasses of graphs.** Most graphs discussed in the literature are subclasses of the graphs considered here. In addition, the global Markov property defined in the next section specializes to the independence structures previously discussed. Exceptions include *MC graphs* [15] and *ribbonless graphs* [27]. However, any independence structure represented by an MC graph or a ribbonless graph can also be represented by a summary graph or an ancestral graph [29], which are also covered in this paper.

Although we do not set any constraints on the class of graphs with four types of edges for the purpose of defining a global Markov property in Section 3, the most general class of graphs for which we explicitly define a pairwise Markov property in Section 5 is the class of *chain mixed graphs* (CMGs) [28]. CMGs are graphs without dotted lines and semi-directed cycles, hence *reciprocal graphs* as in [14]

are not CMGs. CMGs may have multiple edges of all types except a combination of arrows and lines or arrows in opposite directions as such combinations would constitute semi-directed cycles. The graph in Figure 1(a) is an example of graph with four types of edges, and the graph in Figure 1(b) is not a CMG because of the semi-directed cycle  $\langle h \rightarrow p \text{ --- } q \text{ --- } h \rangle$ .

It is helpful to classify subclasses of graphs into three categories: *basic graphs*, *chain graphs* and *mixed graphs*, as briefly described below.

*Basic graphs.* These are graphs that only contain one type of edge; they include *undirected graphs* (UGs), containing only lines; *bidirected graphs* (BGs), containing only bidirected edges; *dotted line graphs* (DGs), containing only dotted lines; and *directed acyclic graphs* (DAGs), containing only arrows without any directed cycle. Clearly, a graph without arrows has no semi-directed cycles, and a semi-directed cycle in a graph with only arrows is a directed cycle. Note that [2, 7, 12, 38] use the terms *concentration graphs* and *covariance graphs* for UGs and BGs, referring to their independence interpretation associated with covariance and concentration matrices for Gaussian graphical models. DGs have not been studied specifically; as we shall see, any independence structure associated with a DG is Markov equivalent to the corresponding UG, where dotted lines are replaced by lines. DAGs have in particular been useful to describe causal Markov relations; see, for example, [10, 13, 17, 22, 31].

*Chain graphs.* A *chain graph* (CG) is a graph with the two following properties: (1) if we remove all arrows, all connected components of the resulting graph—called *chain components*—contain one type of edge only; (2) if we replace every chain component by a node then the resulting graph is a DAG. DAGs, UGs, DGs and BGs are all instances of chain graphs. For a DAG, all chain components are singletons, and for a chain graph without arrows, the chain components are simply the connected components of the graph.

If all chain components contain lines, the chain graph is an *undirected chain graph* (UCG) (here associated with the LWF Markov property); if all contain arcs, it is a *bidirected chain graph* (BCG) (here associated with the multivariate regression chain graph Markov property); and if all contain dotted lines, it is a *dotted line chain graph* (DCG) (here associated with the AMP Markov property). For example, in Figure 2(a) the graph is a chain graph with chain components  $\tau_1 = \{l, j, k\}$ ,  $\tau_2 = \{h, q\}$  and  $\tau_3 = \{p\}$ , but in Figure 2(c) the graph is not a chain graph because of the semi-directed cycle  $\langle h, k, q, h \rangle$ .

*Regression graphs* [42] are chain graphs consisting of lines and arcs (although dashed undirected edges have previously been used instead of arcs in the literature), where there is no arrowhead pointing to nodes that are endpoints of lines.

*Mixed graphs.* Marginalization and conditioning in DCGs (studied in [23]) lead to *marginal AMP graphs* (MAMPs); in our formulation, where we use dotted

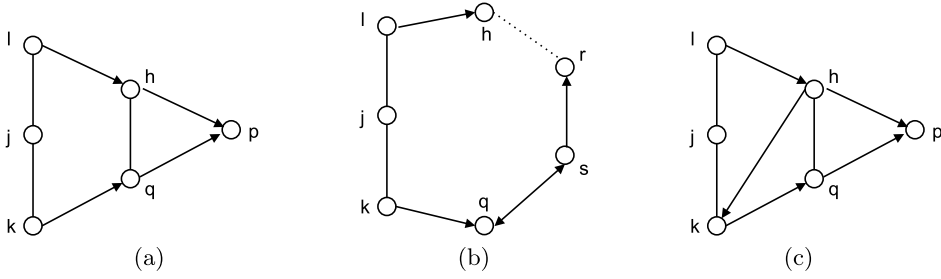


FIG. 2. (a) An undirected chain graph. (b) A chain graph with chain components of different types. (c) A graph that is not a chain graph as  $(h, k, q, h)$  is semi-directed cycle in this graph.

lines in place of full lines, MAMPs are graphs without solid lines that satisfy three additional conditions:

1.  $G$  has no *quasi-directed cycles* in the sense it has no walk  $\langle i = i_0, i_1, \dots, i_n = i \rangle$  containing at least one arrow and every arrow  $i_q i_{q+1}$  is pointing from  $i_q$  to  $i_{q+1}$ ;
2.  $G$  has no cycles composed of dotted lines and one arc;
3. If  $i \dots j \dots k$  and  $j \longleftrightarrow l$  for some  $l$ , then  $i \dots k$ .

Graphs discussed here also contain different types of *mixed graphs*, a term previously used to denote graphs with lines, arrows and arcs. These were introduced to describe independence structures obtained by marginalization and conditioning in DAG independence models; see, for example, [27] for a general discussion of this issue. Examples are *summary graphs* (SGs) [37], *ancestral graphs* (AGs) [26] and *acyclic directed mixed graphs* (ADMGs) [25, 32]. Summary graphs are CMGs that have no arrowhead pointing to nodes that are endpoints of lines. Ancestral graphs satisfy in addition that there are no arcs with one endpoint being an ancestor of the other endpoint. Note that in many papers about summary graphs, dashed undirected edges have been used in place of bidirected edges.

ADMGs are summary graphs without lines. *Alternative ADMGs* (AADMGs) were defined in [24], where arcs in ADMGs were replaced by dotted lines with our notation, although lines were used in the original definition.

CMGs are also mixed graphs, and originally defined in [28] in order to describe independence structures obtained by marginalization and conditioning in chain graph independence models. *Antieral graphs* (AnGs) were also defined in [28] for the same purpose, and they are CMGs in which an endpoint of an arc cannot be an anterior of the other endpoint.

The diagram in Figure 3 illustrates the hierarchy of subclasses of graphs with four types of edges. Below we shall provide a unified separation criterion for all graphs with four types of edges, and thus the associated independence models share the same hierarchy. The diagram is to be read transitively in the sense that, for example, BGs are also AGs, since the class of BGs form a subclass of BCGs,



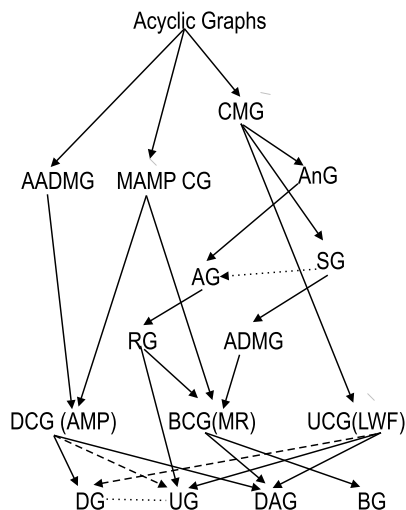


FIG. 3. The hierarchy of graphs with four types of edges and their independence models.

which again form a subclass of AGs; thus we omit the corresponding arrow from AG to BG.

The dashed arrow from DCG to UG indicates that although UGs are not DCGs, their associated independence models contain all independence models given by UGs and similarly for the dashed arrow from UG to DG. The dotted arrow from SG to AG indicates that although AG is a subclass of SG, their associated independence models are the same. The dotted link between UG and DG indicates that the associated independence models are the same. These facts will be demonstrated in the next section.

**3. Graphical independence models.** Graphs are used to encode independence structures for graphical models; in this section we shall demonstrate how this can be done.

**3.1. Independence models and compositional graphoids.** An *independence model*  $\mathcal{I}$  over a finite set  $V$  is a set of triples  $\langle A, B | C \rangle$  (called *independence statements*), where  $A$ ,  $B$  and  $C$  are disjoint subsets of  $V$ ;  $C$  may be empty, but  $\langle \emptyset, B | C \rangle$  and  $\langle A, \emptyset | C \rangle$  are always included in  $\mathcal{I}$ . The independence statement  $\langle A, B | C \rangle$  is read as “ $A$  is independent of  $B$  given  $C$ ”. Independence models may have a probabilistic interpretation (see Section 3.4 for details) but this need not necessarily be the case. Similarly, not all independence models can be easily represented by graphs. For further discussion on general independence models, see [35].

An independence model  $\mathcal{I}$  over a set  $V$  is a *semi-graphoid* if it satisfies the four following properties for disjoint subsets  $A$ ,  $B$ ,  $C$  and  $D$  of  $V$ :

- (S1)  $\langle A, B \mid C \rangle \in \mathcal{J}$  if and only if  $\langle B, A \mid C \rangle \in \mathcal{J}$  (*symmetry*);  
 (S2) if  $\langle A, B \cup D \mid C \rangle \in \mathcal{J}$ , then  $\langle A, B \mid C \rangle \in \mathcal{J}$  and  $\langle A, D \mid C \rangle \in \mathcal{J}$  (*decomposition*);  
 (S3) if  $\langle A, B \cup D \mid C \rangle \in \mathcal{J}$ , then  $\langle A, B \mid C \cup D \rangle \in \mathcal{J}$  and  $\langle A, D \mid C \cup B \rangle \in \mathcal{J}$  (*weak union*);  
 (S4)  $\langle A, B \mid C \cup D \rangle \in \mathcal{J}$  and  $\langle A, D \mid C \rangle \in \mathcal{J}$  if and only if  $\langle A, B \cup D \mid C \rangle \in \mathcal{J}$  (*contraction*).

A semi-graphoid for which the reverse implication of the weak union property holds is said to be a *graphoid*; that is, it also satisfies

- (S5) if  $\langle A, B \mid C \cup D \rangle \in \mathcal{J}$  and  $\langle A, D \mid C \cup B \rangle \in \mathcal{J}$  then  $\langle A, B \cup D \mid C \rangle \in \mathcal{J}$  (*intersection*).

Furthermore, a graphoid or semi-graphoid for which the reverse implication of the decomposition property holds is said to be *compositional*, that is, it also satisfies

- (S6) if  $\langle A, B \mid C \rangle \in \mathcal{J}$  and  $\langle A, D \mid C \rangle \in \mathcal{J}$  then  $\langle A, B \cup D \mid C \rangle \in \mathcal{J}$  (*composition*).

**3.2. Independence models induced by graphs.** The notion of separation is fundamental for using graphs to represent models of independence. For a simple, undirected graph, separation has a direct intuitive meaning, so that a set  $A$  of nodes is separated from a set  $B$  by a set  $C$  if all walks from  $A$  to  $B$  intersect  $C$ . Notice that simple separation in an undirected graph will trivially satisfy all of the properties (S1)–(S6) above, and hence compositional graphoids are abstractions of independence models given by separation in undirected graphs. For more general graphs, separation may be more subtle, to be elaborated below.

We say that a walk  $\omega$  in a graph is *connecting* given  $C$  if all collider sections of  $\omega$  intersect  $C$  and all noncollider sections are disjoint from  $C$ . For pairwise disjoint subsets  $\langle A, B, C \rangle$ , we say that  $A$  and  $B$  are *separated* by  $C$  if there are no connecting walks between  $A$  and  $B$  given  $C$ , and we use the notation  $A \perp B \mid C$ . The set  $C$  is called an  $(A, B)$ -*separator*.

The notion of separation above is a generalization of the *c-separation* for UCGs as defined in [34, 36]. The idea of using walks to simplify the separation theory was proposed by [15], who showed that, for DAGs, this notion of separation was identical to *d-separation* [22].

For example, in the graph of Figure 4,  $j \perp h \mid \{k, l\}$  and  $j \perp h \mid \{k, p\}$  do not hold. The former can be seen by looking at the connecting walk  $\langle j, k, l, r, q, h \rangle$ , where the only node  $k$  and the node  $l$  of the collider sections  $\langle k \rangle$  and  $\langle l, r, q \rangle$  are in the potential separator set  $\{k, l\}$ . The latter can be seen by looking at the connecting walk  $\langle j, k, l, p, l, r, q, h \rangle$ , where the noncollider sections  $\langle l \rangle$  and  $\langle l, r, q \rangle$  are outside  $\{k, p\}$ , but collider sections (nodes)  $\langle k \rangle$  and  $\langle p \rangle$  are inside  $\{k, p\}$ . However, for example,  $j \perp h \mid l$  and  $j \perp h \mid k$  since, in the former case, collider section  $\langle k \rangle$  is

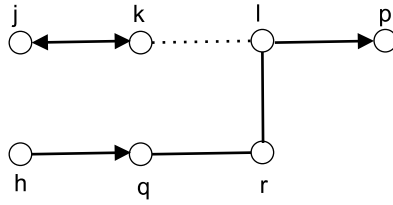


FIG. 4. Illustration of separation in a graph  $G$ ; it holds that  $j \not\perp h \mid \{k, l\}$  and  $j \not\perp h \mid \{k, p\}$ , but  $j \perp h \mid l$  and  $j \perp h \mid k$ .

blocking all the walks and, in the latter case, one of the collider sections  $\langle l, r, q \rangle$  or  $\langle p \rangle$  is blocking any walk.

A graph  $G$  induces an independence model  $\mathcal{I}(G)$  by separation, letting  $\langle A, B \mid C \rangle \in \mathcal{I}(G) \iff A \perp B \mid C$ . It turns out that any independence model defined in this way shares the six fundamental properties of undirected graph separation. More precisely, we have the following.

**THEOREM 1.** *For any graph  $G$ , the independence model  $\mathcal{I}(G)$  is a compositional graphoid.*

**PROOF.** Let  $G = (V, E)$ , and consider disjoint subsets  $A, B, C$  and  $D$  of  $V$ . We verify each of the six properties separately.

(1) *Symmetry:* If  $A \perp B \mid C$ , then  $B \perp A \mid C$ : If there is no connecting walk between  $A$  and  $B$  given  $C$ , then there is no connecting walk between  $B$  and  $A$  given  $C$ .

(2) *Decomposition:* If  $A \perp (B \cup D) \mid C$ , then  $A \perp D \mid C$ : If there is no connecting walk between  $A$  and  $B \cup D$  given  $C$ , then there is *a fortiori* no connecting walk between  $A$  and  $D \subseteq (B \cup D)$  given  $C$ .

(3) *Weak union:* If  $A \perp (B \cup D) \mid C$ , then  $A \perp B \mid (C \cup D)$ : Using decomposition (2) yields  $A \perp D \mid C$  and  $A \perp B \mid C$ . Suppose, for contradiction, that there exists a connecting walk  $\omega$  between  $A$  and  $B$  given  $C \cup D$ . If there is no collider section on  $\omega$ , then there is a connecting walk between  $A$  and  $B$  given  $C$ , a contradiction. On  $\omega$ , all collider sections must have a node in  $(C \cup D)$ . If all collider sections have a node in  $C$ , then there is a connecting walk between  $A$  and  $B$  given  $C$ , again a contradiction. Hence consider first the collider section  $\rho$  nearest  $A$  on  $\omega$  that only has nodes in  $D$  on  $\omega$ ; next, consider the closest node  $i$  to  $A$  on  $\rho$  that is in  $D$ . The subwalk between  $A$  and  $i$  then contradicts  $A \perp B \cup D \mid C$ .

(4) *Contraction:* If  $A \perp B \mid C$  and  $A \perp D \mid (B \cup C)$ , then  $A \perp (B \cup D) \mid C$ : Suppose, for contradiction, that there exists a connecting walk between  $A$  and  $B \cup D$  given  $C$ . Consider a shortest walk (i.e., a walk with fewest number of edges) of this type and call it  $\omega$ . The walk  $\omega$  is either between  $A$  and  $B$  or between  $A$  and  $D$ . The walk  $\omega$  being between  $A$  and  $B$  contradicts  $A \perp B \mid C$ . Therefore,  $\omega$  is between  $A$  and  $D$ . In addition, since all collider sections on  $\omega$  have a node in  $C$  and

$A \perp D \mid (B \cup C)$ , a noncollider section of  $\omega$  must exist that has a node in  $B \cup C$  and, therefore, in  $B$ . This contradicts the fact that  $\omega$  is a shortest connecting walk between  $A$  and  $B \cup D$  given  $C$ .

(5) *Intersection*: If  $A \perp B \mid (C \cup D)$  and  $A \perp D \mid (C \cup B)$ , then  $A \perp (B \cup D) \mid C$ : Suppose, for contradiction, that there exists a connecting walk between  $A$  and  $B \cup D$  given  $C$ . Consider a shortest walk of this type and call it  $\omega$ . The walk  $\omega$  is either between  $A$  and  $B$  or between  $A$  and  $D$ . Because of symmetry between  $B$  and  $D$  in the formulation, it is enough to suppose that  $\omega$  is between  $A$  and  $B$ . Since all collider sections on  $\omega$  have a node in  $C$  and  $A \perp B \mid (C \cup D)$ , a noncollider section of  $\omega$  must exist that has a node in  $D \cup C$  and, therefore, in  $D$ . This contradicts the fact that  $\omega$  is a shortest connecting walk between  $A$  and  $B \cup D$  given  $C$ .

(6) *Composition*: If  $A \perp B \mid C$  and  $A \perp D \mid C$ , then  $A \perp (B \cup D) \mid C$ : Suppose, for contradiction, that there exist connecting walks between  $A$  and  $B \cup D$  given  $C$ . Consider a walk of this type and call it  $\omega$ . The walk  $\omega$  is either between  $A$  and  $B$  or between  $A$  and  $D$ . Because of symmetry between  $B$  and  $D$  in the formula it is enough to suppose that  $\omega$  is between  $A$  and  $B$ . But this contradicts  $A \perp B \mid C$ .  $\square$

This theorem implies that we can focus on establishing conditional independence for pairs of nodes, formulated in the corollary below.

**COROLLARY 1.** *For a graph  $G$  and disjoint subsets of nodes  $A$ ,  $B$  and  $C$ , it holds that  $A \perp B \mid C$  if and only if  $i \perp j \mid C$  for every pair of nodes  $i \in A$  and  $j \in B$ .*

**PROOF.** The result follows from the fact that  $\perp$  satisfies decomposition and composition.  $\square$

**3.3. Relation to other separation criteria.** Four different types of independence models have previously been associated with chain graphs. These are known as the LWF Markov property, defined by [18] and later studied by, for example, [9, 36]; the AMP Markov property, defined and studied by [1], and the multivariate regression (MR) Markov property, introduced by [2] and studied, for example, by [20]; in addition, Drton [6] briefly considered a type III chain graph Markov property which we are not further considering here.

Traditionally, these have been formulated using undirected chain graphs but with different separation criteria. In contrast, here we use a single notion of separation and the different independence models appear by varying the type of chain graph. In particular, the LWF Markov property corresponds to UCGs, the MR Markov property to BCGs and the AMP Markov property to DCGs, as we shall see below.

Table 1 gives an overview of different types of colliders used in the various independence models associated with chain graphs.

For summary graphs and their subclasses, [29] showed that the unifying separation concept was that of *m-separation*, defined as follows. A path  $\pi$  is *m-connecting* given  $C$  if all collider nodes on  $\pi$  intersect  $An(C)$  and all non-collider

TABLE 1  
Colliders for different chain graph independence models

MR:	$\rightarrow \circ \leftarrow$	$\rightarrow \circ \leftrightarrow$	$\leftrightarrow \circ \leftrightarrow$
AMP:	$\rightarrow \circ \leftarrow$		$\rightarrow \circ \dots$
LWF:		$\rightarrow \circ \text{---} \dots \text{---} \circ \leftarrow$	

nodes on  $\pi$  are disjoint from  $C$ . Notice that  $m$ -separation considers nodes, but the fact that there is no arrowhead pointing to a node that is endpoint of a line in a summary graph implies that every collider section of any walk consists of a single node. For pairwise disjoint subsets  $\langle A, B, C \rangle$ ,  $A$  and  $B$  are  $m$ -separated by  $C$  if there are no  $m$ -connecting paths between  $A$  and  $B$  given  $C$ , and we use the notation  $A \perp_m B \mid C$  to indicate this. The following lemma establishes that for summary graphs (and all subclasses of these),  $m$ -separation is equivalent to the separation we have defined here. The idea is similar to that employed in [15].

LEMMA 2. *Suppose that  $G$  is a summary graph. Then*

$$A \perp B \mid C \iff A \perp_m B \mid C.$$

PROOF. We need to show that for  $i, j \notin C$ , there is a connecting walk between  $i$  and  $j$  if and only if there is an  $m$ -connecting path between  $i$  and  $j$  given  $C$ . If there is an  $m$ -connecting path  $\pi$  between  $i$  and  $j$ , then there exists a connecting walk between  $i$  and  $j$  by taking  $\pi$  and add the possible directed path from a collider node  $k$  on  $\pi$  to  $c \in C$  and its reverse from  $c$  to  $k$ .

Thus suppose that there is a connecting walk  $\omega$  between  $i$  and  $j$ . Since there are no arrowheads pointing to nodes that are endpoints of lines, all collider sections on  $\omega$  are single nodes; and hence we can talk of collider nodes instead of sections. Consider the walk between  $i$  and  $j$  obtained from  $\omega$  by replacing any subwalk of type  $\langle l, \rho', l \rangle$  (for a subwalk  $\rho'$ ) by a single node subwalk  $\langle l \rangle$ . First of all, it is clear that the resulting walk is a path. Denote this path by  $\pi$ . We show that an  $m$ -connecting path can be constructed from  $\pi$ :

It is not possible that a node that occurs (at least once) as a collider on  $\omega$  and occurs also as a member of a noncollider section on  $\pi$ : If  $k$  is a collider node on  $\omega$ , then it is in  $C$ . This means that there is an arrowhead at  $k$  on all tripaths with inner node  $k$  on  $\omega$ . Hence, regardless of which two edges of  $\omega$  with endpoint  $k$  are on  $\pi$ , the corresponding tripath remains collider.

Therefore, all noncollider nodes on  $\pi$  are outside  $C$ . If all collider nodes are in  $C$ , then we are done. Thus, suppose that there is a collider node  $k$  (on collider tripath  $\langle k_0, k, k_1 \rangle$ ) on  $\pi$  that is not in  $C$ . This means that, on  $\omega$ ,  $k$  is always within a noncollider section. Consider an edge  $kr_0$  on  $\omega$  that is a part of the subwalk  $\langle k_0, k, r_0 \rangle$  of  $\omega$ , and notice that this edge is not on  $\pi$ . The edge  $kr_0$  is not a line as otherwise there is an arrowhead pointing to an endpoint of a line. As the edge  $kr_0$

itself has no arrowhead at  $k$ , it must be an arrow from  $k$  to  $r_0$ . Following through  $\omega$  from  $r_0$ , inductively, we have three cases: (1) There exists a directed cycle, which is impossible. (2)  $k$  is an ancestor of a collider node  $r$ : We have that  $r \in C$ , and hence  $k$  is an ancestor of  $C$ . (3)  $k$  is an ancestor of  $i$  or  $j$ : Without loss of generality, assume that  $k \in \text{an}(j)$ . In this case, we modify  $\pi$  by replacing the subwalk between  $k$  and  $j$  by a directed path from  $k$  to  $j$ . Notice that no node on this path is in  $C$ . This completes the proof.  $\square$

For MAMPs, [23] provides a generalization of the  $p$ -separation [19] for AMP chain graphs. In the language and notation of this paper, it is defined as follows: A path  $\pi$  is  $z$ -connecting given  $C$  ( $z$  is our notation) for MAMPs if every collider node on  $\pi$  is in  $\text{An}(C)$  and every noncollider node  $k$  is outside  $C$  unless there is a subpath of  $\pi$ ,  $i \cdots k \cdots j$  such that  $\text{sp}(k) \neq \emptyset$  or  $\text{pa}(k) \setminus C \neq \emptyset$ . We say that  $A$  and  $B$  are  $z$ -separated given  $C$ , and write  $A \perp_z B \mid C$ , if there is no  $z$ -connecting path between  $A$  and  $B$  given  $C$ .

LEMMA 3. *Suppose that  $G$  is a MAMP. Then*

$$A \perp B \mid C \iff A \perp_z B \mid C.$$

PROOF. We need to show that for  $i, j \notin C$ , there is a connecting walk between  $i$  and  $j$  if and only if there is a  $z$ -connecting path between  $i$  and  $j$  given  $C$ . If there is a  $z$ -connecting path  $\pi$  between  $i$  and  $j$ , we may construct a connecting walk between  $i$  and  $j$  by modifying  $\pi$  as follows: (1) for a collider node  $k \in \text{an}(C)$ , add a directed path from  $k$  to  $c \in C$  and its reverse from  $c$  to  $k$ ; (2) for a noncollider node  $k \in C$  within  $i \cdots k \cdots j$  in  $\pi$  (see the definition of  $z$ -separation), we distinguish two cases: if  $\text{sp}(k) \neq \emptyset$  then one has  $i \cdots j$  by the definition of MAMP and one can shorten the tripath  $i \cdots k \cdots j$  on  $\pi$ ; if  $\text{sp}(k) = \emptyset$  but  $l \in \text{pa}(k) \setminus C$  exists then add the  $kl$  edge and its reverse to  $\pi$ .

Thus suppose that there is a connecting walk  $\omega$  between  $i$  and  $j$ . Since there are no lines, all sections on  $\omega$  are single nodes; and hence we can talk of collider and noncollider nodes instead of sections. Similar to Lemma 2, consider the walk between  $i$  and  $j$  obtained from  $\omega$ , and whenever there is a node  $l$  with repeated occurrence in  $\omega$ , replace the cycle from  $l$  to  $l$  in  $\omega$  by a single occurrence of  $l$ . The resulting walk is a path, denoted by  $\pi$ . We show that  $z$ -connecting path can be constructed from  $\pi$ :

The only case where a node  $k$  is a collider node on  $\omega$  and it turns into a non-collider node on  $\pi$  is when  $k$  is the inner node of the tripath  $h \cdots k \cdots l$  on  $\pi$ . Therefore, all noncollider nodes on  $\pi$  are outside  $C$  unless this mentioned case occurs. However, in this case either  $\text{sp}(k) \neq \emptyset$  or  $\text{pa}(k) \setminus C \neq \emptyset$ , which ensures that the condition of the definition of a  $z$ -connecting path is still satisfied.

If all collider nodes are in  $C$ , then we are done. Thus suppose that there is a collider node  $k$  (on collider tripath  $\langle k_0, k, k_1 \rangle$ ) on  $\pi$  that is not in  $C$ . This means

that, on  $\omega$ ,  $k$  is always a noncollider node. There is an arrowhead at  $k$  on at least one of the  $k_0k$  or the  $kk_1$  edges. Without loss of generality, assume that it is the  $k_0k$  edge. Consider an edge  $kr_0$  on  $\omega$  that is a part of the subwalk  $\langle k_0, k, r_0 \rangle$  of  $\omega$ , and notice that this edge is not on  $\pi$ . As the edge  $kr_0$  itself has no arrowhead at  $k$  and is not a dotted line, it must be an arrow from  $k$  to  $r_0$ . Following through  $\omega$  from  $r_0$ , inductively, we have three cases: (1) There exists a directed cycle, which is impossible. (2)  $k$  is an ancestor of a collider node  $r$ : We have that  $r \in C$ , and hence  $k$  is an ancestor of  $C$ . (3)  $k$  is an ancestor of  $i$  or  $j$ : Without loss of generality, assume that  $k \in \text{an}(j)$ . In this case, we modify  $\pi$  by replacing the subwalk between  $k$  and  $j$  by a directed path from  $k$  to  $j$ . Notice that no node on this path is in  $C$ . This completes the proof.  $\square$

We are now ready to show that our concept of separation unifies the independence models discussed.

**THEOREM 2.** *Independence models generated by separation in graphs with four types of edges are identical to the independence models associated with the subclasses in Figure 3.*

**PROOF.** It is shown in [29] that  $m$ -separation, as defined above, unifies independence models for SGs and subclasses thereof and by Lemma 2  $m$ -separation is equivalent to our separation. The separation criterion in [28] for CMGs is identical to the separation given here when there are no dotted lines in the graph. Hence, the independence models generated by our separation criterion unifies independence models for all the subclasses of CMGs. Lemma 3 shows that, dotted lines replacing lines in Peña's separation criterion, it becomes identical to ours. For AADMGs, Criterion 2 defined as the global Markov property in [24] is trivially a special case of the separation defined here. Therefore, our criterion unifies independence models in all subclasses of graphs.  $\square$

Notice that most of the associated classes of independence models presented in the diagram of Figure 3 are distinct; exceptions are AGs and SGs, which are alternative representations of the same class of independence models, and the same holds for DGs and UGs. In addition, it can be seen from Table 1 that, for every type of chain graph, one different type of symmetric edge is needed since each of them forms different colliders; hence, the unification for the general class of graphs with four types of edges is not achieved by graphs with three types of edges.

**3.4. Probabilistic independence models and the global Markov property.** Consider a set  $V$  and a collection of random variables  $(X_\alpha)_{\alpha \in V}$  with state spaces  $\mathcal{X}_\alpha$ ,  $\alpha \in V$  and joint distribution  $P$ . We let  $X_A = (X_v)_{v \in A}$ , etc. for each subset  $A$  of  $V$ . For disjoint subsets  $A$ ,  $B$  and  $C$  of  $V$ , we use the short notation  $A \perp\!\!\!\perp B \mid C$

to denote that  $X_A$  is *conditionally independent of*  $X_B$  *given*  $X_C$  [4, 16], that is, that for any measurable  $\Omega \subseteq \mathcal{X}_A$  and  $P$ -almost all  $x_B$  and  $x_C$ ,

$$P(X_A \in \Omega \mid X_B = x_B, X_C = x_C) = P(X_A \in \Omega \mid X_C = x_C).$$

We can now induce an independence model  $\mathcal{J}(P)$  by letting

$$\langle A, B \mid C \rangle \in \mathcal{J}(P) \quad \text{if and only if} \quad A \perp\!\!\!\perp B \mid C \quad \text{w.r.t. } P.$$

We note that for a probabilistic independence model  $\mathcal{J}(P)$ , the marginal independence model to a set  $D = V \setminus M$  is the independence model generated by the marginal distribution. More formally, we define the *marginal independence model* over a subset of the node set  $M$  as follows:

$$\alpha(\mathcal{J}, M) = \{ \langle A, B \mid C \rangle : \langle A, B \mid C \rangle \in \mathcal{J} \text{ and } (A \cup B \cup C) \cap M = \emptyset \},$$

which is defined over  $V \setminus M$ .

LEMMA 4. *Let  $\mathcal{J}(P)$  be a probabilistic independence model; its marginal independence model is the independence model generated by the marginal distribution, that is, for  $D = V \setminus M$  we have  $\alpha(\mathcal{J}(P), M) = \mathcal{J}(P_D)$ .*

PROOF. This is immediate.  $\square$

For a graph  $G = (V, E)$ , an independence model  $\mathcal{J}$  defined over  $V$  satisfies the *global Markov property* w.r.t. a graph  $G$ , if for disjoint subsets  $A, B$  and  $C$  of  $V$  it holds that

$$A \perp\!\!\!\perp B \mid C \implies \langle A, B \mid C \rangle \in \mathcal{J}.$$

If  $\mathcal{J}(P)$  satisfies the global Markov property w.r.t. a graph  $G$ , we also say that  $P$  is *Markov* w.r.t.  $G$ . We say that an independence model  $\mathcal{J}$  is *probabilistic* if there is a distribution  $P$  such that  $\mathcal{J} = \mathcal{J}(P)$ . We then also say that  $P$  is *faithful* to  $\mathcal{J}$ . If  $P$  is faithful to  $\mathcal{J}(G)$  for a graph  $G$  then we also say that  $P$  is *faithful to*  $G$ . Thus, if  $P$  is faithful to  $G$  it is also Markov w.r.t.  $G$ .

Probabilistic independence models are always semi-graphoids [22], whereas the converse is not necessarily true; see [33]. If, for example,  $P$  has strictly positive density, the induced independence model is always a graphoid; see, for example, Proposition 3.1 in [16]. If the distribution  $P$  is a regular multivariate Gaussian distribution,  $\mathcal{J}(P)$  is a compositional graphoid; for example, see [35].

Probabilistic independence models with positive densities are not in general compositional; this only holds for special types of multivariate distributions such as, for example, Gaussian distributions and the symmetric binary distributions used in [41]. However, the following statement implies that it is not uncommon for a probabilistic independence model to satisfy composition.

PROPOSITION 1. *If there is a graph  $G$  to which  $P$  is faithful, then  $\mathcal{J}(P)$  is a compositional graphoid.*

PROOF. The result follows from Theorem 1 since then  $\mathcal{J}(P) = \mathcal{J}(G)$ .  $\square$



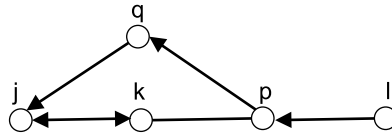


FIG. 5. A nonmaximal CMG.

**4. Maximality for graphs.** A graph  $G$  is called *maximal* if adding an edge between any two nonadjacent nodes in  $G$  changes the independence model  $\mathcal{I}(G)$ . Notice that in [29] the nonadjacency condition was incorrectly omitted.

Conditions 2 and 3, which MAMPs satisfy (provided in Section 2.2) ensure that MAMPs are maximal; see [23]. However, graphs are not maximal in general. For example, there exist nonmaximal ancestral and summary graphs [26, 29]; see also Figure 5 for an example of a graph that is neither a summary graph (hence it is not ancestral) nor maximal: this CMG induces no independence statements of the form  $j \perp l \mid C$  for any choice of  $C$ : if we condition on  $k$  or  $p$  or both, the path  $\langle j, k, p, l \rangle$  is connecting since  $k \rightarrow p$  is a collider section; conditioning on  $q$  makes the walk  $\langle j, k, p, q, p, l \rangle$  a connecting walk, and if we do not condition on anything, the walk  $\langle j, q, p, l \rangle$  is connecting.

The notion of maximality is important for pairwise Markov properties, to be discussed in the next section. For a nonmaximal ancestral or summary graph, one can obtain a maximal ancestral or summary graph with the same induced independence model by adding edges to the original graph [26, 29]. As we shall show below, this is also true for general CMGs, but it is not generally the case for graphs containing dotted lines or directed cycles. Figure 6 displays two small *nonmaximizeable* graphs, where the graph in (a) contains a directed cycle.

For example, in the directed graph of Figure 6(a), in order to make the graph maximal, one must connect  $h$  and  $k$ , and similarly  $l$  and  $j$ . Now notice that in the original graph it holds that  $h \perp l$  and  $h \perp l \mid \{j, k\}$ . However, after introducing new  $hk$  and  $lj$  edges, regardless of what type of edge we add, one of  $h \perp l$  or  $h \perp l \mid \{j, k\}$  does not hold.

To characterise maximal CMGs, we need the following notion: A walk  $\omega$  is a *primitive inducing walk* between  $i$  and  $j$  ( $i \neq j$ ) if and only if it is an  $ij$  edge or  $\omega = \langle i, q_1, q_2, \dots, q_p, j \rangle$  where for every  $n$ ,  $1 \leq n \leq p$ , it holds that:

- (i) all inner sections of  $\omega$  are colliders;

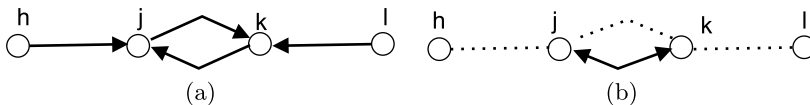


FIG. 6. Two nonmaximal graphs that cannot be modified to be maximal by adding edges without changing the independence model.

- (ii) endpoint sections of  $\omega$  are single elements;
- (iii)  $q_n \in \text{ant}(\{i, j\})$ .

This definition is an extension of the notion of a primitive inducing path as defined for ancestral graphs in [26]. For example, in Figure 5,  $\langle j, k, p, l \rangle$  is a primitive inducing walk. Next, we need the following lemmas.

**LEMMA 5.** *In a CMG, inner nodes of a walk  $\omega$  between  $i$  and  $j$  that are on a noncollider section are either in  $\text{ant}(i) \cup \text{ant}(j)$  or antérieurs of a collider section on  $\omega$ .*

**PROOF.** Let  $k = i_m$  be an inner node of  $\omega$  and on a noncollider section on a walk  $\omega = \langle i = i_0, i_1, \dots, i_n = j \rangle$  in a CMG  $G$ . Then from at least one side (say from  $i_{m-1}$ ), there is no arrowhead on  $\omega$  pointing to the section containing  $k$ . By moving towards  $i$  on the path as long as  $i_p$ ,  $1 \leq p \leq m-1$ , is on a noncollider section on the walk, we obtain that  $k \in \text{ant}(i_{p-1})$ . This implies that if no  $i_p$  is on a collider section then  $k \in \text{ant}(i)$ .  $\square$

**LEMMA 6.** *For nodes  $i$  and  $j$  in a CMG that are not connected by any primitive inducing walks (and hence  $i \approx j$ ), it holds that  $i \perp j \mid \text{ant}(\{i, j\})$ .*

**PROOF.** Suppose that there is a connecting walk  $\varpi$  between  $i$  and  $j$  given  $\text{ant}(\{i, j\})$ .

If  $i$  or  $j$  are on a noncollider inner section  $\rho$  on  $\varpi$ , then  $\rho$  is contained in  $\{i, j\}$  since otherwise any other node in  $\rho$  would be in  $\text{ant}(\{i, j\})$ , which is impossible. Then  $\rho$  contains either only  $i$  or only  $j$  since  $ij$  is not an edge in the graph. Thus,  $\rho$  is either single  $i$  or single  $j$ . In such a case, remove the cycle between  $i$  and  $i$  (or between  $j$  and  $j$ ), which is a subwalk of  $\varpi$ . Repeat this process until there are no such noncollider sections. Denote the resulting walk by  $\omega$ . We shall show that  $\omega$  is primitive inducing:

(i) If, for contradiction, there is a node  $k$  on an inner noncollider section of  $\omega$  then, by Lemma 5,  $k$  is either in  $\text{ant}(i) \cup \text{ant}(j)$  or it is an anterior of nodes of a collider section on  $\omega$ , but since  $\omega$  is connecting given  $\text{ant}(\{i, j\})$ , collider sections intersect  $\text{ant}(\{i, j\})$ , and hence are in  $\text{ant}(\{i, j\})$  themselves. (Hence,  $k \notin \{i, j\}$ .) Now,  $k \in \text{ant}(\{i, j\})$  contradicts the fact that  $\omega$  is connecting given  $\text{ant}(\{i, j\})$ .

(ii) Unless  $\omega$  is a line, endpoint sections of  $\omega$  are single elements since they are noncollider on  $\omega$  and, if not single elements, their members, excluding  $i$  or  $j$ , are in  $\text{ant}(\{i, j\})$ , which is impossible.

(iii) This condition is clear since all inner nodes are in collider sections and consequently (except for possibly  $i$  or  $j$ ) in  $\text{ant}(i) \cup \text{ant}(j)$ .  $\square$

**LEMMA 7.** *The only primitive inducing walk between  $i$  and  $j$  without arrowheads at its endpoints (i.e.,  $i$  and  $j$ ) is the line  $ij$ .*

PROOF. Consider such a walk  $\omega$ : Suppose, for contradiction, that there are other nodes other than  $i, j$  on  $\omega$ , and assume that  $q_1$  is the node adjacent to the endpoint  $i$  on  $\omega$  (i.e., there is  $i \rightarrow q_1$ ). Notice that  $q_1 \neq j$  since otherwise  $i \rightarrow j$ ,  $j \rightarrow q_p$ , for some  $p$ , and  $q_p \in \text{Ant}(\{i, j\})$  lead to a contradiction.

Then the lack of semi-directed cycles implies that  $q_1 \in \text{ant}(j)$ , and hence there is another node  $q_2$  on  $\omega$ . Similarly, for  $q_p$  adjacent to  $j$  on  $\omega$ ,  $q_p \in \text{ant}(i)$ . But we may then construct a semi-directed cycle by taking the  $iq_1$  edge, the anterior path from  $q_1$  to  $j$ , the  $jq_p$  edge, and the anterior path from  $q_p$  back to  $i$ , a contradiction.  $\square$

Next, we say that two walks  $\omega_1$  and  $\omega_2$  (including edges) between  $i$  and  $j$  are *endpoint-identical* if there is an arrowhead pointing to the endpoint section containing  $i$  in  $\omega_1$  if and only if there is an arrowhead pointing to the endpoint section containing  $i$  in  $\omega_2$  and similarly for  $j$ . For example, the paths  $i \rightarrow j$ ,  $i \rightarrow k \rightarrow l \leftrightarrow j$ , and  $i \rightarrow k \leftrightarrow l \rightarrow j$  are all endpoint-identical as they have an arrowhead pointing to the section containing  $j$  but no arrowhead pointing to the section containing  $i$  on the paths, but they are not endpoint-identical to  $i \rightarrow k \leftrightarrow j$ . We then have the following.

LEMMA 8. *Let  $G$  be a CMG with the node set  $V$ . If there is a primitive inducing walk  $\omega$  between  $i$  and  $j$  in  $G$ , and  $C \subseteq V \setminus \{i, j\}$ , then a connecting walk between  $i$  and  $j$  given  $C$  exists which is endpoint identical to  $\omega$ .*

PROOF. We denote the sections of the primitive inducing walk  $\omega$  by  $\langle i = \tau_0, \tau_1, \dots, \tau_{s-1}, \tau_s = j \rangle$  and note that if a section  $\tau$  intersects  $\text{ant}(A)$  for any set  $A$ , it holds that  $\tau \subseteq \text{ant}(A)$ . By Lemma 7, it is enough to consider two cases:

*Case 1.* There is an arrowhead at  $j$  and no arrowhead at  $i$  on  $\omega$ : First notice that the edge  $iq_1$  is an arrow from  $i$  to  $q_1$ . We construct an endpoint-identical connecting walk  $\varpi$  given  $C$  between  $i$  and  $j$ . We start from  $i$  and move towards  $j$  on  $\omega$  via  $iq_1$  where  $q_1 \in \tau_1$ . As long as along  $\varpi$ , a section  $\tau_n$ ,  $1 \leq n \leq s-1$ , intersects  $\text{Ant}(C)$ , we do the following: If  $\tau_n \cap C \neq \emptyset$ , then we let  $\varpi$  move to  $\tau_{n+1}$ . If  $\tau_n \cap C = \emptyset$ , but  $\tau_n \subseteq \text{ant}(C)$  then we let  $\varpi$  move from  $\tau_n$  to  $C$  via an anterior path and back to  $\tau_n$  by reversing this path, subsequently continuing to  $\tau_{n+1}$  using the corresponding edge from  $\omega$ .

So suppose that possibly  $\varpi$  reaches a section  $\tau_m$  not intersecting  $\text{Ant}(C)$ . Note that  $\tau_m$  cannot only contain  $i$  since otherwise it intersects  $\text{Ant}(C)$  (through  $\tau_1$ ). If  $\tau_m$  only contains  $j$ , then we already have a connecting walk. Hence, the only case that is left is when there is a  $k \in \tau_m \setminus \{i, j\}$  such that  $k \in \text{ant}(\{i, j\})$ . If  $k \in \text{ant}(i)$ , then notice that  $k$  is an anterior of  $C$  through  $i$  and  $\tau_1$ , which is impossible. Thus  $k \in \text{ant}(j)$  with no nodes on the anterior path in  $C$ . We can now complete  $\varpi$  by letting it move to  $j$  via this anterior path.

Notice that  $\varpi$  is endpoint-identical to  $\omega$  since both have an arrowhead at  $j$  and no arrowhead at  $i$ .

*Case 2.* There is an arrowhead at  $j$  and an arrowhead at  $i$  on  $\omega$ : We follow the same method as in Case 1 to construct  $\varpi$ . The only difference is that  $k \in \tau_m \setminus \{i, j\}$  can be in  $\text{ant}(i)$  without being an anterior of  $C$ . (In fact,  $k$  and  $i$  may be on the same section on  $\omega$ .) In this case, we entirely replace the already constructed part of  $\varpi$  by the reverse of the anterior path from  $k$  to  $i$  (which is from  $i$  to  $k$ ), and let  $\varpi$  proceed to  $\tau_{m+1}$ .

Again it is clear that the constructed  $\varpi$  and  $\omega$  have an arrowhead at  $j$ . If  $k$  and  $i$  are not in the same section or are not connected by an undirected path, then it is clear that there is an arrowhead at  $i$ , which is a single-node section on  $\varpi$ . If  $k$  and  $i$  are in the same section or are connected by an undirected path, then there is an arrowhead at the endpoint section of  $\varpi$  that contains  $i$ .  $\square$

Next, in Theorem 3 we give a necessary and sufficient condition for a CMG to be maximal. The analogous result for ancestral graphs was proved in Theorem 4.2 of [26].

**THEOREM 3.** *A CMG  $G$  is maximal if and only if  $G$  does not contain any primitive inducing walks between nonadjacent nodes.*

**PROOF.**  $(\Rightarrow)$  Let  $\omega = \langle i = i_0, i_1, \dots, i_n = j \rangle$  be a primitive inducing walk between nonadjacent nodes  $i$  and  $j$ . By Lemma 8, there is therefore an endpoint-identical connecting walk  $\omega'$  between  $i$  and  $j$  given any choice of  $C$ ; thus, there is clearly no separation of form  $i \perp j \mid C$ . Let us add an endpoint-identical  $ij$  edge to  $G$ . If a separation  $A \perp B \mid C'$  is destroyed, then the edge  $ij$  is a part of the connecting walk  $\omega''$  given  $C'$  between  $A$  and  $B$ . Now by replacing  $ij$  by  $\omega'$  on  $\omega''$ , we clearly obtain a walk in  $G$  that is connecting given  $C'$ . This implies that adding  $ij$  does not change  $\mathcal{J}(G)$ ; hence,  $G$  is not maximal.

$(\Leftarrow)$  By letting  $C = \text{ant}(\{i, j\})$  for every nonadjacent pair of nodes  $i$  and  $j$  and using Lemma 6, we conclude that for every missing edge there is an independence statement in  $\mathcal{J}(G)$ . This implies that  $G$  is maximal.  $\square$

It now follows that for maximal graphs, every missing edge corresponds to a pairwise conditional independence statement in  $\mathcal{J}(G)$ .

**COROLLARY 2.** *A CMG  $G$  is maximal if and only if every missing edge in  $G$  corresponds to a pairwise conditional independence statement in  $\mathcal{J}(G)$ .*

**PROOF.**  $(\Leftarrow)$  is clear.  $(\Rightarrow)$  follows from Theorem 3 and Lemma 6.  $\square$

Also, we have the following corollary.

**COROLLARY 3.** *If  $G$  is a nonmaximal CMG, then it can be made maximal by adding edges without changing its independence model.*

PROOF. We begin with a nonmaximal CMG  $G$ , and show that we can “close” all the primitive inducing walks in order to obtain a maximal graph with the same induced independence model. For every primitive inducing walk  $\omega$  between  $i$  and  $j$  where  $i \approx j$  in  $G$ , add an  $ij$  edge that is endpoint-identical to  $\omega$  if an edge of the same type does not already exist.

First, we show that the resulting graph is a CMG: It is enough to show that an added edge does not generate a semi-directed cycle. By Lemma 7, the added edge is either an arrow or an arc. Since arcs are not on a semi-directed cycle, adding an arc would not generate a semi-directed cycle. Thus, suppose that the added edge is an arrow from  $i$  to  $j$ . Notice that the adjacent node  $q_1$  to  $i$  on the primitive inducing walk is in  $\text{ant}(j)$  and the  $iq_1$  edge is an arrow from  $i$  to  $q_1$ . Hence, if, for contradiction, the added  $ij$  arrow generates a semi-directed cycle, a semi-directed cycle already existed in the original graph, where  $ij$  is replaced by the anterior walk that consists of the  $iq_1$  arrow and the anterior walk from  $q_1$  to  $j$ . This is a contradiction.

Now, since the resulting graph does not contain any primitive inducing walks between nonadjacent nodes, it is maximal. In addition, by Lemma 8, there is a connecting walk between  $i$  and  $j$ , which is endpoint-identical to the primitive inducing walk. One can replace the endpoint-identical  $ij$  edge to this walk in any connecting walk in  $G$  that contains  $\omega$  as a subwalk.  $\square$

For example, in Figure 5,  $\langle j, k, p, l \rangle$  was a primitive inducing walk; hence this graph was not maximal. We may then add the edge  $l \longrightarrow j$  and it becomes maximal.

## 5. Pairwise Markov properties for chain mixed graphs.

5.1. *A pairwise Markov property.* It is possible to consider a general pairwise Markov property for specific subclasses of graphs with four types of edges (that actually have the four types) by including the results of [23, 24], which define pairwise Markov properties for marginal AMP chain graphs and alternative ADMGs and show the equivalence of pairwise and global Markov properties for such graphs. However, such a unification would be technically complex. Hence, we henceforth focus on CMGs; thus, the considerations here concerning pairwise Markov properties do not cover AMP chain graphs.

A pairwise Markov property provides independence statements for non-adjacent pairs of nodes in the graph. For maximal graphs, any nonadjacent nodes  $i$  and  $j$  are independent given some set  $C$ , but a pairwise Markov property yields a specific choice of  $S = S(i, j)$  for every nonadjacent pair  $i, j$ . The choice we provide here for any CMG immediately extends the choice in [29]. We show that for a maximal CMG, this pairwise Markov property is equivalent to the global Markov property for compositional graphoid independence models; in other words, the pairwise statements combined with the compositional graphoid axioms generate

the full independence model. The maximality is critical for the pairwise statements to hold, as discussed above.

An independence model  $\mathcal{J}$  defined over  $V$  satisfies the *pairwise Markov property* (P) w.r.t. a CMG  $G$  if for every pair of nodes  $i$  and  $j$  with  $i \approx j$  it holds that

$$(P): \quad \langle i, j \mid \text{ant}(\{i, j\}) \rangle \in \mathcal{J}.$$

The pairwise Markov property simplifies for specific subclasses of graphs. For connected UGs, we have  $\text{ant}(\{i, j\}) = V \setminus \{i, j\}$ , and hence the standard pairwise Markov property appears; and for BGs we have  $\text{ant}(\{i, j\}) = \emptyset$ , so the property is identical to pairwise independence of nonadjacent nodes. For SGs and AGs (which include DAGs), a semi-direction preserving path is of the form  $\circ \text{---} \cdots \circ \text{---} \circ \longrightarrow \cdots \circ \longrightarrow$ , hence the anterior path [and consequently (P)] specializes to those in [29] and [26], respectively.

Strictly speaking, the unification only contains “connected” UGs. It is not possible to extend the unification to all UGs and at the same time keep the pairwise Markov properties defined in the literature for other classes under any unified pairwise Markov property: In principle, it is fine to add nodes that are not in the connected component(s) of  $i$  and  $j$  to the conditioning set in any pairwise Markov property. However, although the well-known pairwise Markov property for UGs contains all such nodes, the known pairwise Markov properties for other classes do not.

**5.2. Equivalence of pairwise and global Markov properties.** Before establishing the main result of this section, we need several lemmas. We shall need to consider marginalization of independence models and use that it preserves the compositional graphoid property, shown in Lemma 8 of [29].

**LEMMA 9.** *Let  $\mathcal{J}$  be a compositional graphoid over a set  $V$  and  $M$  a subset of  $V$ . It then holds that the marginal independence model  $\alpha(\mathcal{J}, M)$  is also a compositional graphoid.*

Moreover, we have the following.

**LEMMA 10.** *Let  $\mathcal{J} = \mathcal{J}(G)$  be the independence model induced by a CMG  $G$  and  $M \subseteq V$ . If  $D = V \setminus M$  is an anterior set, the marginal model is determined by the induced subgraph  $G[D]$ :*

$$\alpha(\mathcal{J}(G), M) = \mathcal{J}(G[D]).$$

**PROOF.** We need to show that for  $\{i, j\} \cup C \subseteq D$  we have that  $i \perp j \mid C$  if and only if this is true in the induced subgraph  $G[D]$ . Clearly, if a connecting walk between  $i$  and  $j$  runs entirely within  $D$  it also connects in  $G$ . Assume for

contradiction that there is a connecting walk which has a node  $k$  outside  $D$  and consider an excursion on the walk that leaves  $D$  at  $i^*$ , reaches  $k$  and reenters into  $D$  at  $j^*$ . Since the walk is connecting, there are no collider sections on this excursion, and thus it follows from Lemma 5 that  $k$  is either anterior to  $i^*$  or to  $j^*$ , which contradicts the fact that  $D$  is an anterior set.  $\square$

The following important lemma and its corollary imply that for any non-adjacent pair  $i \approx j$  in a maximal CMG we can always find an  $(i, j)$ -separator  $C$  with  $C \subseteq \text{ant}(\{i, j\})$ .

**LEMMA 11.** *For a pair of distinct nodes  $i$  and  $j$  and a subset of the node set  $C$  in a maximal CMG, if  $i \perp j \mid C$  for  $C \setminus \text{ant}(\{i, j\}) \neq \emptyset$ , then there is a node  $l \notin \text{ant}(\{i, j\})$  in  $C$  such that  $i \perp j \mid C \setminus \{l\}$ .*

**PROOF.** Let  $l' \in C \setminus \text{ant}(\{i, j\})$  be arbitrary. If there is an  $l'' \in C \setminus \text{ant}(\{i, j\})$  so that  $l' \in \text{ant}(l'')$  but  $l'' \notin \text{ant}(l')$ , then replace  $l'$  by  $l''$ , and repeat this process until it terminates, which is ensured by the transitivity of semi-directed walks and the lack of semi-directed cycles in the CMG. Call the final node  $l$ . Thus, if  $l \in \text{ant}(\tilde{l})$  for  $\tilde{l} \in C \setminus \text{ant}(\{i, j\})$  then we also have that  $\tilde{l} \in \text{ant}(l)$ . The lack of semi-directed cycles implies that this is equivalent to  $l$  and  $\tilde{l}$  being connected by lines.

We now claim that  $i \perp j \mid C \setminus \{l\}$ . Suppose, for contradiction, that there is a connecting walk  $\omega$  between  $i$  and  $j$  given  $C \setminus \{l\}$ . If  $l$  is not on  $\omega$ , then  $\omega$  is also connecting given  $C$ . In addition, we have that  $l$  is on a non-collider section  $\rho$  on  $\omega$ . There is no arrowhead at  $\rho$  from at least one side of the section, say from the  $i$  side. We move towards  $i$  on  $\omega$  and denote the corresponding subwalk of  $\omega$  by  $\omega' = \langle l = l_0, l_1, \dots, l_m = i \rangle$ . As long as  $l_p, 1 \leq p \leq m - 1$ , is on a noncollider section on  $\omega'$ , we obtain that there is a semi-directed walk from  $l$  to  $l_p$ . This implies that if no  $l_p$  is on a collider section then there is an anterior walk from  $l$  to  $i$ , which is impossible.

Therefore, by moving towards  $i$  from  $l$ , we first reach an  $\tilde{l}$  on  $\omega'$  that lies on a collider section and is in  $C \setminus \{l\}$ . Transitivity of anterior walks and the fact that there is no anterior walk from  $l$  to  $i$  or  $j$  now imply that there is no anterior walk from  $\tilde{l}$  to  $i$  or  $j$ . The construction of  $l$  implies that  $l$  and  $\tilde{l}$  are on the same section, and hence  $l$  is not on a noncollider section on  $\omega$ , a contradiction. Hence, we conclude that  $i \perp j \mid C \setminus \{l\}$ .  $\square$

**COROLLARY 4.** *For a pair of nodes  $i$  and  $j$  and a subset  $C$  of the node set in a maximal CMG, if  $i \perp j \mid C$ , then  $i \perp j \mid C \cap \text{ant}(\{i, j\})$ .*

**PROOF.** Lemma 11 implies that we can repeatedly remove single nodes in  $C \setminus \text{ant}(\{i, j\})$  and preserve separation to obtain that  $i \perp j \mid C \cap \text{ant}(\{i, j\})$ . This concludes the proof.  $\square$



A direct implication of Lemma 6 and Theorem 3 establishes that the induced independence model  $\mathcal{J}(G)$  for a maximal CMG  $G$  satisfies the pairwise Markov property (P).

**PROPOSITION 2.** *If  $i \approx j$  are nonadjacent nodes in a maximal CMG  $G$ , it holds that  $i \perp j \mid \text{ant}(\{i, j\})$ .*

Finally, we are ready to show the main result of this section.

**THEOREM 4.** *Let  $G$  be a maximal CMG. If an independence model  $\mathcal{J}$  over the node set of  $G$  is a compositional graphoid, then  $\mathcal{J}$  satisfies the pairwise Markov property (P) w.r.t.  $G$  if and only if it satisfies the global Markov property w.r.t.  $G$ .*

**PROOF.** That the global Markov property implies the pairwise property (P) follows directly from Proposition 2.

Now suppose that  $\mathcal{J}$  satisfies the pairwise Markov property (P) and compositional graphoid axioms. For subsets  $A$ ,  $B$  and  $C$  of the node set of  $G$ , we must show that  $A \perp B \mid C$  implies  $\langle A, B \mid C \rangle \in \mathcal{J}$ . Since the separation satisfies decomposition and  $\mathcal{J}$  satisfies composition, it is sufficient to show this when  $A$  and  $B$  are singletons, that is, that  $i \perp j \mid C$  implies  $\langle i, j \mid C \rangle \in \mathcal{J}$ .

We establish the result in two main parts. In part I, we consider the case with  $C \subseteq \text{ant}(\{i, j\})$  and in part II we consider the general case.

*Part I.* Suppose that  $C \subseteq \text{ant}(\{i, j\})$ . We use induction on the number of nodes of the graph. The induction base for a graph with two nodes is trivial. Thus, suppose that the conclusion holds for all graphs with fewer than  $n$  nodes and assume that  $G$  has  $n$  nodes.

Suppose there is an anterior set  $D$  such that  $M = V \setminus D \neq \emptyset$  and  $\{i\} \cup \{j\} \cup C \subseteq D$ . The marginal independence model  $\alpha(\mathcal{J}, M)$  clearly also satisfies the pairwise Markov property w.r.t.  $G[D]$ , and hence the inductive assumption together with Lemmas 9 and 10 yields  $\langle \{i\}, \{j\} \mid C \rangle \in \mathcal{J}$ .

So suppose that this is not the case, and hence  $V = \text{Ant}(\{i, j\})$ . We establish the conclusion by reverse induction on  $|C|$ : For the base, we have  $C = V \setminus \{i, j\} = \text{ant}(\{i, j\})$  and the result follows directly from the pairwise Markov property.

For the inductive step, consider a node  $h \notin C$ . We want to show that there are not simultaneously connecting walks between  $h$  and  $i$  and  $h$  and  $j$ : Suppose, for contradiction, there are connecting walks  $\omega_1 = \langle i, i_1, \dots, i_n, h \rangle$  and  $\omega_2 = \langle h, j_m, j_{m-1}, \dots, j_0 = j \rangle$  given  $C$ . If, on the walk  $\langle \omega_1, \omega_2 \rangle$ , the node  $h$  is on a noncollider section then so is  $h$  on both  $\omega_1$  and  $\omega_2$ , and hence  $i$  and  $j$  are connected given  $C$ , a contradiction. Thus, we need only consider the case where  $h$  is on a collider section on  $\langle \omega_1, \omega_2 \rangle$ . However, we know that  $h \in \text{ant}(i)$  or  $h \in \text{ant}(j)$ . Because of symmetry between  $i$  and  $j$  suppose that  $h \in \text{ant}(j)$ , and denote the anterior path from  $h$  to  $j$  by  $\omega_3$ . Notice that the section containing  $h$  on  $\omega_1$  is non-collider, and hence all members are outside  $C$ . Now, if no node on  $\omega_3$  is in  $C$  then



$\langle \omega_1, \omega_3 \rangle$  is a connecting walk between  $i$  and  $j$  a contradiction; and if there is a node  $k$  on  $\omega_3$  is in  $C$  then  $\langle \omega_1, \omega_4, \omega_4', \omega_2 \rangle$  is a connecting walk between  $i$  and  $j$ , where  $\omega_4$  is the subwalk of  $\omega_3$  between  $h$  and  $k$  and  $\omega_4'$  is  $\omega_4$  in reverse direction, a contradiction again. We conclude that, given  $C$ ,  $h$  is not connected to both  $i$  and  $j$ .

By symmetry, suppose that  $i \perp h \mid C$ . We also have that  $i \perp j \mid C$ . Since  $\mathcal{J}(G)$  is a compositional graphoid (Theorem 1) the composition property gives that  $i \perp \{j, h\} \mid C$ . By weak union for  $\perp$ , we obtain  $i \perp j \mid \{h\} \cup C$  and  $i \perp h \mid \{j\} \cup C$ . By the induction hypothesis, we obtain  $\langle i, j \mid \{h\} \cup C \rangle \in \mathcal{J}$  and  $\langle i, h \mid \{j\} \cup C \rangle \in \mathcal{J}$ . By intersection, we get  $\langle i, \{j, h\} \mid C \rangle \in \mathcal{J}$ . By decomposition, we finally obtain  $\langle i, j \mid C \rangle \in \mathcal{J}$ .

*Part II.* We now prove the result in the general case by induction on  $|C|$ . The base, that is, the case that  $|C| = 0$ , follows from part I. To prove the inductive step, we can assume that  $C \not\subseteq \text{ant}(\{i, j\})$ , since otherwise part I implies the result.

By Lemma 11, since  $C \not\subseteq \text{ant}(\{i, j\})$ , there is a node  $l \in C$  such that  $i \perp j \mid C \setminus \{l\}$ . We now have that either  $i \perp l \mid C \setminus \{l\}$  or  $j \perp l \mid C \setminus \{l\}$  since otherwise there is a connecting walk between  $i$  and  $j$  given  $C \setminus \{l\}$  in the case that  $l$  is on a noncollider section or given  $C$  in the case that  $l$  is on a collider section. Because of symmetry, suppose that  $i \perp l \mid C \setminus \{l\}$ . By the induction hypothesis, we have  $\langle i, j \mid C \setminus \{l\} \rangle \in \mathcal{J}$  and  $\langle i, l \mid C \setminus \{l\} \rangle \in \mathcal{J}$ . By the composition property, we get  $\langle i, \{j, l\} \mid C \setminus \{l\} \rangle \in \mathcal{J}$ . The weak union property implies  $\langle i, j \mid C \rangle \in \mathcal{J}$ .  $\square$

If we specialize Theorem 4 to the most common case of probabilistic independence models, we get the following.

**COROLLARY 5.** *Let  $G$  be a maximal CMG. A probabilistic independence model that satisfies the intersection and composition axioms satisfies the pairwise Markov property (P) w.r.t.  $G$  if and only if it satisfies the global Markov property w.r.t.  $G$ .*

The theorem states that the intersection and composition properties are sufficient for equivalence of pairwise and global Markov properties. Notice that they are also necessary since, for example, for the simple subclass of DAGs they are also necessary; see Section 6.3 of [29].

**5.3. Alternative pairwise Markov properties.** There are typically many other valid choices of the separating sets  $C(i, j)$  defining the pairwise Markov properties; see, for example, [23]. In general, a pairwise Markov property (P\*) has the form

$$(P^*): \quad i \approx j \quad \Rightarrow \quad \langle i, j \mid C(i, j) \rangle \in \mathcal{J},$$

where  $C(i, j)$  is an  $(i, j)$ -separator in  $G$  for every  $(i, j)$ . The question then is what are the possible choices of such separator systems which would ensure that

these separations form a ‘basis’ for the independence model  $\mathcal{I}$  in the sense that all conditional independences in  $\mathcal{I}$  can be derived from  $(P^*)$  using the compositional graphoid axioms. The example below shows that not all choices of separator systems are possible.

EXAMPLE 1. Consider the independence model  $\mathcal{I}$  over  $V = \{1, 2, 3, 4, 5\}$  containing the statements

$$\langle 1, 3 \mid 2 \rangle, \langle 1, 4 \mid 3 \rangle, \langle 1, 5 \mid 4 \rangle, \langle 2, 4 \mid 1, 3, 5 \rangle, \langle 2, 5 \mid 3 \rangle, \langle 3, 5 \mid 1, 2, 4 \rangle$$

as well as their symmetric counterparts and all independence statements of the form  $\langle A, \emptyset \mid B \rangle$  or  $\langle \emptyset, A \mid B \rangle$ . This independence model is easily seen to satisfy the compositional graphoid axioms. In addition, if we let  $G$  be the graph  $1 \text{ --- } 2 \text{ --- } 3 \text{ --- } 4 \text{ --- } 5$ , each of the conditioning sets for statements of the form  $\langle i, j \mid C(i, j) \rangle$  in  $\mathcal{I}$  are indeed  $(i, j)$ -separators in  $G$ . Thus,  $\mathcal{I}$  satisfies  $(P^*)$  w.r.t. the graph  $G$ , but clearly it does not satisfy the global Markov property w.r.t.  $G$ .

Note that the independence model in Example 1 may not be probabilistically representable. It is unclear to us whether the pairwise statements in  $(P^*)$  for any system  $C(i, j)$  of  $(i, j)$ -separators in an undirected graph, say, is sufficient to generate all independence statements of the form  $\langle A, B \mid C \rangle$  for a *probabilistic* compositional graphoid. The standard probabilistic counterexamples [21] involving the pairwise Markov properties are not compositional graphoids, hence are not counterexamples in this context.

For the subclass of regression graphs, four different pairwise Markov properties were defined in [30], which are all equivalent to the global Markov property and to each other under compositional graphoid axioms.

**6. Summary and conclusion.** In this paper, we used a similar approach to that of [29] to unify Markov properties for most classes of graphs in the literature of graphical models. The general idea is that for any of the three standard interpretations of the chain graph Markov property, (LWF, AMP and multivariate regression), we use one type of edge in the unifying class of graphs and then use a single separation criterion which is a natural generalization of  $c$ -separation as defined in [36].

Unifying an equivalent pairwise Markov property seems very technical when including the AMP chain graphs, hence we restricted ourselves to prove the equivalence of pairwise and global Markov properties for the class of maximal CMGs, which includes connected chain graphs with the LWF interpretation as well as maximal summary graphs (and consequently maximal ancestral graphs). In order to include the class of AMP chain graphs or its generalizations for the unification of the pairwise Markov property, excluding certain “directed cycles” from the class

of graphs with four types of edges (similar to the exclusion of semi-directed cycles in mixed graphs) is necessary.

It was seen in this paper that, under compositional graphoid axioms, the system of pairwise independence statements constituting the pairwise Markov property, can act as a generating class for all independence statements given by the global Markov property. Typically, there are many other systems of pairwise statements that may act as a generating class for the global Markov property. The point given here is that there is a unified choice of these statements for the case of CMGs.

The two important independence models are induced by graphs and probability distributions. Establishing the pairwise Markov property for the independence model induced by graphs suffices for establishing the global Markov property as it is always a compositional graphoid. This is not always the case for the independence model induced by any probability distribution as the intersection and composition properties may not hold for such distributions.

**Acknowledgements.** The authors are very grateful to the anonymous reviewers for careful, detailed and helpful comments.

## REFERENCES

- [1] ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (2001). Alternative Markov properties for chain graphs. *Scand. J. Stat.* **28** 33–85. DOI:[10.1111/1467-9469.00224](#). [MR1844349](#)
- [2] COX, D. R. and WERMUTH, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.* **8** 204–218, 247–283. [MR1243593](#)
- [3] DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539. [MR0568718](#)
- [4] DAWID, A. P. (1979). Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc. Ser. B* **41** 1–31.
- [5] DIDELEZ, V. (2008). Graphical models for marked point processes based on local independence. *J. Roy. Statist. Soc. Ser. B* **70** 245–264.
- [6] DRTON, M. (2009). Discrete chain graph models. *Bernoulli* **15** 736–753. DOI:[10.3150/08-BEJ172](#). [MR2555197](#)
- [7] DRTON, M. and RICHARDSON, T. S. (2008). Binary models for marginal independence. *J. Roy. Statist. Soc. Ser. B* **41** 287–309.
- [8] EICHLER, M. (2007). Granger causality and path diagrams for multivariate time series. *J. Econometrics* **137** 334–353.
- [9] FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353.
- [10] GEIGER, D., VERMA, T. S. and PEARL, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534.
- [11] GIBBS, W. (1902). *Elementary Principles of Statistical Mechanics*. Yale Univ. Press, New Haven, CT.
- [12] KAUFMANN, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Stat.* **23** 105–116. [MR1380485](#)
- [13] KIIVERI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Aust. Math. Soc. A* **36** 30–52.
- [14] KOSTER, J. T. A. (1996). Markov properties of nonrecursive causal models. *Ann. Statist.* **24** 2148–2177. [MR1421166](#)

- [15] KOSTER, J. T. A. (2002). Marginalizing and conditioning in graphical models. *Bernoulli* **8** 817–840. [MR1963663](#)
- [16] LAURITZEN, S. L. (1996). *Graphical Models. Oxford Statistical Science Series* **17**. Clarendon Press, Oxford. [MR1419991](#)
- [17] LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Roy. Statist. Soc. Ser. B* **50** 157–224. [MR0964177](#)
- [18] LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57. [MR0981437](#)
- [19] LEVITZ, M., PERLMAN, M. D. and MADIGAN, D. (2001). Separation and completeness properties for amp chain graph Markov models. *Ann. Statist.* **29** 1751–1784. [MR1891745](#)
- [20] MARCHETTI, G. M. and LUPPARELLI, M. (2011). Chain graph models of multivariate regression type for categorical data. *Bernoulli* **17** 827–844. [MR2817607](#)
- [21] MATÚŠ, F. (1992). On equivalence of Markov properties over undirected graphs. *J. Appl. Probab.* **29** 745–749. [MR1174448](#)
- [22] PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo, CA.
- [23] PEÑA, J. M. (2014). Marginal AMP chain graphs. *Internat. J. Approx. Reason.* **55** 1185–1206. [MR3197803](#)
- [24] PEÑA, J. M. (2016). Alternative Markov and causal properties for acyclic directed mixed graphs. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence, UAI 2016, June 25–29, 2016, Jersey City, NJ*.
- [25] RICHARDSON, T. (2003). Markov properties for acyclic directed mixed graphs. *Scand. J. Stat.* **30** 145–157.
- [26] RICHARDSON, T. S. and SPIRITES, P. (2002). Ancestral graph Markov models. *Ann. Statist.* **30** 962–1030. [MR1926166](#)
- [27] SADEGHI, K. (2013). Stable mixed graphs. *Bernoulli* **19** 2330–2358. [MR3160556](#)
- [28] SADEGHI, K. (2016). Marginalization and conditioning for LWF chain graphs. *Ann. Statist.* **44** 1792–1816. [MR3519941](#)
- [29] SADEGHI, K. and LAURITZEN, S. L. (2014). Markov properties for mixed graphs. *Bernoulli* **20** 676–696. [MR3178514](#)
- [30] SADEGHI, K. and WERMUTH, N. (2016). Pairwise Markov properties for regression graphs. *Stat* **5** 286–294.
- [31] SPIRITES, P., GLYMOUR, C. and SCHEINES, R. (2000). *Causation, Prediction, and Search*, 2nd ed. MIT Press, Cambridge, MA.
- [32] SPIRITES, P., RICHARDSON, T. and MEEK, C. (1997). The dimensionality of mixed ancestral graphs. Technical Report No. CMU-PHIL-83, Philosophy Dept., Carnegie Mellon Univ., Pittsburgh, PA.
- [33] STUDENÝ, M. (1989). Multiinformation and the problem of characterization of conditional independence relations. *Probl. Control Inf. Theory* **18** 3–16.
- [34] STUDENÝ, M. (1998). Bayesian networks from the point of view of chain graphs. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* 496–503. Morgan Kaufmann, San Francisco, CA.
- [35] STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. Springer, London, United Kingdom.
- [36] STUDENÝ, M. and BOUCKAERT, R. R. (1998). On chain graph models for description of conditional independence structures. *Ann. Statist.* **26** 1434–1495. [MR1647685](#)
- [37] WERMUTH, N. (2011). Probability distributions with summary graph structure. *Bernoulli* **17** 845–879.

- [38] WERMUTH, N. and COX, D. R. (1998). On association models defined over independence graphs. *Bernoulli* **4** 477–495. [MR1679794](#)
- [39] WERMUTH, N., COX, D. R. and PEARL, J. (1994). Explanation for multivariate structures derived from univariate recursive regressions. Technical Report No. 94(1), Univ. Mainz, Germany.
- [40] WERMUTH, N. and LAURITZEN, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70** 537–552.
- [41] WERMUTH, N., MARCHETTI, G. M. and COX, D. R. (2009). Triangular systems for symmetric binary variables. *Electron. J. Stat.* **3** 932–955. DOI:[10.1214/09-EJS439](#). [MR2540847](#)
- [42] WERMUTH, N. and SADEGHI, K. (2012). Sequences of regressions and their independences. *TEST* **21** 215–252, 274–279.
- [43] WRIGHT, S. (1921). Correlation and causation. *J. Agric. Res.* **20** 557–585.

DEPARTMENT OF MATHEMATICAL SCIENCES  
UNIVERSITY OF COPENHAGEN  
UNIVERSITETSPARKEN 5  
COPENHAGEN, 2100  
DENMARK  
E-MAIL: [lauritzen@math.ku.dk](mailto:lauritzen@math.ku.dk)

STATISTICAL LABORATORY  
CENTRE FOR MATHEMATICAL SCIENCES  
WILBERFORCE ROAD  
CAMBRIDGE, CB3 0WA  
UNITED KINGDOM  
E-MAIL: [k.sadeghi@statslab.cam.ac.uk](mailto:k.sadeghi@statslab.cam.ac.uk)